B-082E
c.2

Statistics Canada Statistique Canada

Ottawa

REPORT ON THE ANALYSIS OF THE FCC PILOT SURVEY SAMPLE DESIGN

Susan Ingram
Institutional and Agriculture
Survey Methods Division

March, 1981

Report on the Analysis of the FCC Pilot Survey Sample Design

1. Introduction

Statistics Canada was approached in 1979 by the Farm Credit Corporation (F.C.C.) for assistance in designing a survey to determine the structure of agricultural assets and liabilities by province. Specifically, the F.C.C. wanted to compare their borrowers and non-F.C.C. clients with respect to their balance of assets and debts. As the first stage in the development of a full scale survey, a pilot study was conducted in February 1980 in two F.C.C. field districts. One of the ways in which the results of the pilot were analyzed involved evaluating the pilot survey design. The purpose of this report is to document these results and to make recommendations for the sample design of the main survey based on this evaluation.

2. Sample Design of Pilot Survey

The sample design of the pilot survey was formulated by J. Phillips of Institutional and Agriculture Survey Methods Division. For background for this report, a brief description of the design will be provided in this section; however, further details can be found in "Design of FCC Pilot Survey", B. Armstrong, July 1980[1].

Two FCC field districts were chosen for the pilot - Quebec Region 6 and Saskatchewan Region 1. The study population was also restricted to active farms having total sales of $1200 and over (in 1976) since farms with under $1200 sales were considered to be marginal farmers and were not of interest. For each region, FCC files of farmers who had received a loan were matched to active farms on the Farm Register, resulting in the following three groups being formed: (1) Farms which were on both files; (2) Farms which were on the Farm Register but not on the FCC file; and (3) Farms which were on the FCC file but not on the Farm Register.

Farms in groups 1 and 2 (farms on the Farm Register) were stratified according to an asset value obtained from the census where total asset value was defined as the sum of the value of land, the value of machinery and the value of livestock. For farms in group 3, no information was available for stratification. Within each region, then, 15 strata were formed with stratum ranges defined in consultation with the sponsor. There were 7 strata for each of groups 1 and 2 with farms in group 3 forming the remaining stratum. Within each group, all strata with less than 20 farms were combined and all farms in the resulting stratum were included in the sample. All other strata were given an equal allocation of 10 farms per stratum with the sample selected using systematic sampling.

3. Results

3.1 Gain due to stratification

As the first step in evaluating the sample design, the gain in precision relative to simple random sampling was calculated for four key variables included in the pilot survey - total assets, total liabilities, total area and total cultivated area. It should be noted that the total asset value obtained in the survey does not correspond to the asset value used for stratification. The former not only includes the value of livestock, land and machinery (the stratification variable) but also such items as cash, bonds, savings, accounts receivable, crops for sale, etc.

Calculating the efficiency of the stratified design is accomplished by comparing the estimated variance of the mean from the stratified sample with an estimate of the variance of the mean that would have been obtained from a simple random sample. For stratified random sampling, the estimate of the population mean $\bar{Y}$ is

$$\bar{y}_{st} = \frac{\sum_h N_h \bar{y}_h}{N} = \sum_h F_h \bar{y}_h \qquad (3.1)$$

and the estimated variance of $\bar{y}_{st}$ is given by

$$v(\bar{y}_{st}) = \sum_h \frac{F_h^2 s_h^2}{n_h} - \sum_h \frac{F_h s_h^2}{N} \qquad (3.2)$$

where h = subscript denoting the hth stratum ($h = 1, \ldots L$);

N_h = population size of stratum h;

n_h = number of sample units selected from stratum h;

$$N = \sum_h N_h; \quad n = \sum_h n_h;$$

$$F_h = N_h/N; \quad \bar{y}_h = \sum_i y_{hi}/n_h;$$

and $$s_h^2 = \frac{\sum_i y_{hi}^2 - \frac{(\sum_i y_{hi})^2}{n_h}}{n_h - 1}$$

The estimate of what the variance would have been had simple random sampling been employed is given

$$v_{ran} = \frac{N-n}{n(N-1)} \left[\sum_h F_h s_h^2 - \sum_h \frac{F_h s_h^2}{n_h} + \sum_h \frac{F_h^2 s_h^2}{n_h} - \sum_h \frac{F_h s_h^2}{N} + \sum_h F_h \bar{y}_h^2 - (\sum_h F_h \bar{y}_h)^2\right] \qquad (3.3)$$

The above expression is unattractive to compute and can be simplified for large N. Since $(N-1) \doteq N$ and by omitting the fourth term in the bracket since $(1 - \frac{1}{N}) \doteq 1$, we obtain

$$v_{ran} \doteq \frac{N-n}{nN} \left[\sum_h F_h s_h^2 - \sum_h \frac{F_h s_h^2}{n_h} + \sum_h \frac{F_h^2 s_h^2}{n_h} + \sum_h F_h \bar{y}_h^2 - (\sum_h F_h \bar{y}_h)^2\right] \qquad (3.4)$$

The following table contains these estimated variances for each of the key variables in both Saskatchewan Region 1 and Quebec Region 6.

TABLE 1: Relative Efficiency of Stratification

Variable	Quebec Region 6				Saskatchewan Region 1			
	$\bar{y}_{st}$	V_{st}	V_{ran}	V_{st}/V_{ran}	$\bar{y}_{st}$	V_{st}	V_{ran}	V_{st}/V_{ran}
Total assets	306053	907,658,945.0	409,226,299.2	2.22	392229	1,292,001,789.8	714,061,826.7	1.81
Total liabilities	61493	97,887,076.4	60,759,281.0	1.61	43026	33,610,380.6	23,874,123.6	1.41
Total area	126	243.7	125.2	1.95	651	6,999.3	3,053.9	2.29
Total cult. area	107	199.95	93.7	2.13	500	3,664.8	1,909.5	1.92

It is clear that in all cases studied there was a loss in precision in the estimates resulting from the stratification employed in the pilot survey. As it was expected that there would have been a gain due to stratification, the results were next analyzed to determine if the allocation to the strata had been poor.

3.2 Allocation

For each of the four key variables, the optimum allocation given the sample size of the pilot survey was studied. The optimum allocation for a fixed total sample size of n is given by

$$n_h = n \frac{N_h S_h}{\sum_h N_h S_h} \quad \text{with} \qquad (3.5)$$

$$V_{min}(\bar{y}_{st}) = \frac{(\Sigma F_h S_h)^2}{n} - \frac{\Sigma F_h S_h^2}{N} \qquad (3.6)$$

where S_h^2 is the true variance in stratum h.

To determine if the stratum sample sizes employed in the pilot were far from optimum, the optimum allocation was computed separately for each item by substituting the sample variance, s_h^2, in (3.5). The resulting n_h are found in Appendix 1. These results show that the equal allocation employed in the pilot survey design was substantially different from the individual optima. In general, it appears that the sample should have been larger in those strata with low total asset values and relatively smaller samples should have been selected from strata with high total assets.

Since the individual optima differ from each other, some compromise allocation must be found. In the pilot survey the optimum allocations for assets, total area and cultivated area differed only moderately from each other. However, a somewhat different allocation was indicated for liabilities, especially in the Saskatchewan data. Two compromise allocations were investigated--proportional allocation and allocation proportional to $\sqrt{N_h}$, which gives a relatively larger sample to smaller strata. The compromise allocations are also given in the tables in Appendix 1; the table below shows the expected sampling variances (expressed in terms of the coefficient of variation of $\bar{y}_{st}$ as given by the individual optima, proportional allocation and allocation proportion to $\sqrt{N_h}$. The estimated minimum variances are given by (3.6) and the expected variances of the compromise allocations are found by substituting the appropriate n_h's in (3.2).

TABLE 2: Estimated Sampling Variances of $\bar{y}_{st}$

Variable	Quebec Region 6				Saskatchewan Region 1			
	$\bar{y}_{st}$	c.v. (opt.)	c.v. (prop.)	c.v. (prop. $\sqrt{N_h}$)	$\bar{y}_{st}$	c.v. (opt.)	c.v. (prop.)	c.v. (prop. $\sqrt{N_h}$)
al assets	306,053	5.3	5.6	6.1	392,229	5.4	5.8	6.2
al liabilities	61,493	9.5	13.0	10.6	43,026	8.0	9.9	9.7
al area	126	7.0	7.5	7.9	651	7.4	7.9	8.2
al cult. area	107	7.3	8.0	8.5	500	7.4	7.9	8.2

It appears that allocation proportional to $\sqrt{N_h}$ is a good compromise; with this allocation the estimated coefficients of variation of average assets, area and cultivated area increase relative to proportional allocation but the estimated c.v. of average liabilities decreases. However, proportional allocation is only slightly less precise than the individual optima with the exception of average liabilities. Subject to constraints of cost and staff availability imposed on the sample size of the full-scale survey, allocation of the sample proportional to $\sqrt{N_h}$ should be considered. This allocation has the added advantage of somewhat larger sample sizes in the smaller strata which would mean that more FCC clients would be included in the sample. This is advantageous since a comparison between this group and non-FCC clients is a major objective of the proposed survey.

3.3 The Construction of the Strata

The stratum definitions employed in the pilot survey design had been specified in consultation with the FCC and Agriculture Division. As part of the analysis of the pilot results, these definitions were compared with the strata boundaries which would have been used had the method for determining optimum strata developed by V.K. Sethi[2] been used. The exact ranges of each stratum in the pilot design cannot be given here due to rules of confidentiality for the census; however, the optimum stratification given the number of strata of the pilot design can be found in Appendix 2. Consistent with the analysis of the sample allocation, optimum stratification would have resulted in fewer farms with high asset values being included in the sample. For the proposed full-scale survey, Sethi's technique for determining optimum strata will be used.

APPENDIX 1

SAMPLE ALLOCATION: QUEBEC REGION 6

Stratum Description		Stratum I.D.	Pop^{n} Size	Pilot n_h	Opt. n_h (Assets)	Opt. n_h (Liab.)	Opt. n_h (Total Area)	Opt. n_h (Cult. Area)	Prop.to N_h	Prop.to $\sqrt{N_h}$
All farms on Central Register with '76 Census sales ≥$1200	Farms not on FCC file	1	2,333	10	55.0	45.2	48.8	47.1	56.7	31.7
		2	1,272	10	32.0	35.4	35.8	38.9	30.9	23.4
		3	196	10	6.7	4.8	11.4	11.9	4.8	11.0
		4	30	10	2.1	3.1	1.9	1.8	.7	3.6
		5	13	13	1.9	4.8	2.5	1.7	.3	2.4
	Farms on FCC file	6	191	10	3.0	3.0	4.8	3.7	4.7	9.1
		7	513	10	5.2	10.6	6.3	4.7	12.5	14.9
		8	168	10	7.6	4.7	.3	1.8	4.7	8.5
		9	16	16	1.8	2.0	1.9	2.0	.4	2.6
	Farms not on FCC - assets N.A.	10	137	10	2.5	3.5	2.1	1.6	3.3	7.7
	Farms on FCC - assets N.A.	11	9	9	.2	.3	.3	.3	.2	2.0
All farms on FCC file which did not match to a farm on Central Register		12	391	10	10.4	10.7	11.9	12.4	9.5	13.0

SAMPLE ALLOCATION: Saskatchewan Region 1

Stratum Description		Stratum I.D.	Pop^{n} Size	Pilot n_h	Opt. n_h (Assets)	Opt. n_h (Liab.)	Opt. n_h (Total Area)	Opt. n_h (Cult. Area)	Prop.to N_h	Prop.to $\sqrt{N_h}$
All farms on Central Register with '76 Census sales ≥$1200	Farms not on FCC file	1	3,332	10	19.6	52.9	15.4	16.4	33.7	23.9
		2	4,851	10	46.0	10.7	55.0	46.7	49.1	28.8
		3	2,016	10	39.3	19.6	28.3	30.2	20.4	18.6
		4	264	10	2.6	4.9	3.9	3.8	2.7	6.7
		5	73	10	1.4	2.2	1.7	1.9	.7	3.5
		6	57	10	1.8	2.4	2.4	2.3	.6	3.1
	Farms on FCC file	7	323	10	3.4	7.7	2.5	2.9	3.3	7.4
		8	1,364	10	9.3	10.1	11.2	12.9	13.8	15.3
		9	950	10	9.0	11.7	12.7	13.3	9.6	12.8
		10	117	10	1.2	2.0	1.1	1.4	1.2	4.5
		11	20	10	.3	.3	.3	.3	.2	1.9
		12	9	9	.2	.6	.1	.2	.1	1.2
	Farms not on FCC - assets N.A.	13	446	10	4.9	.5	4.0	4.2	4.5	8.7
	Farms on FCC - assets N.A.	14	14	14	.2	.2	.2	.2	.1	1.5
All farms on FCC file which did not match to a farm on Central Register		15	1,294	10	13.8	27.0	14.2	15.6	13.1	14.9

APPENDIX 2

OPTIMUM STRATIFICATION: QUEBEC REGION 6

Stratum Description			Stratum Number
All farms on Central Register with '76 Census sales ≥$1200	Farms not on FCC file	Total Assets:	
		$ 1 - 65,442	1
		65,443 - 113,676	2
		113,677 - 188,163	3
		188,164 - 368,796	4
		> 368,796	5
	Farms on FCC file	Total Assets:	
		$ 1 - 129,681	6
		129,682 - 222,218	7
		222,219 - 419,047	8
		> 419,047	9

OPTIMUM STRATIFICATION: SASKATCHEWAN REGION 1

Stratum Description			Stratum Number
All farms on Central Register with '76 Census sales ≥$1200	Farms not on FCC file	Total Assets:	
		$ 1 - 92,329	1
		92,330 - 163,332	2
		163,333 - 249,559	3
		249,560 - 383,517	4
		383,518 - 739,600	5
		> 739,600	6
	Farms on FCC file	Total Assets:	
		$ 1 - 140,645	7
		140,646 - 215,110	8
		215,111 - 289,223	9
		289,224 - 382,927	10
		382,927 - 565,270	11
		> 565,270	12

REFERENCES

1) Armstrong, B. "Design of the F.C.C. Pilot Survey" (1980)

2) Sethi, V.K. "A Note on Optimum Stratification of Populations for Estimating the Population Means" Australian Journal of Statistics, 5 (1963).